Машинное обучение для Временных Точечных Процессов

Алексей Зайцев

Старший преподаватель, Skoltech

Читает: Владислав Жужель

Аспирант, Skoltech

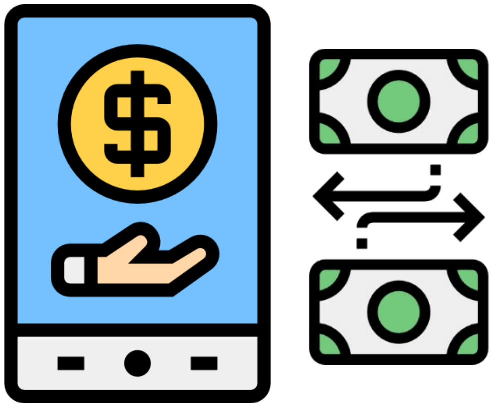
Основано на ICML Tutorial, July 2018 by I. Valera & M.G. Rodriguez

Слайды и ссылки: http://learning.mpi-sws.org/tpp-icml18

Временные точечные процессы (TPPs):
Приложения

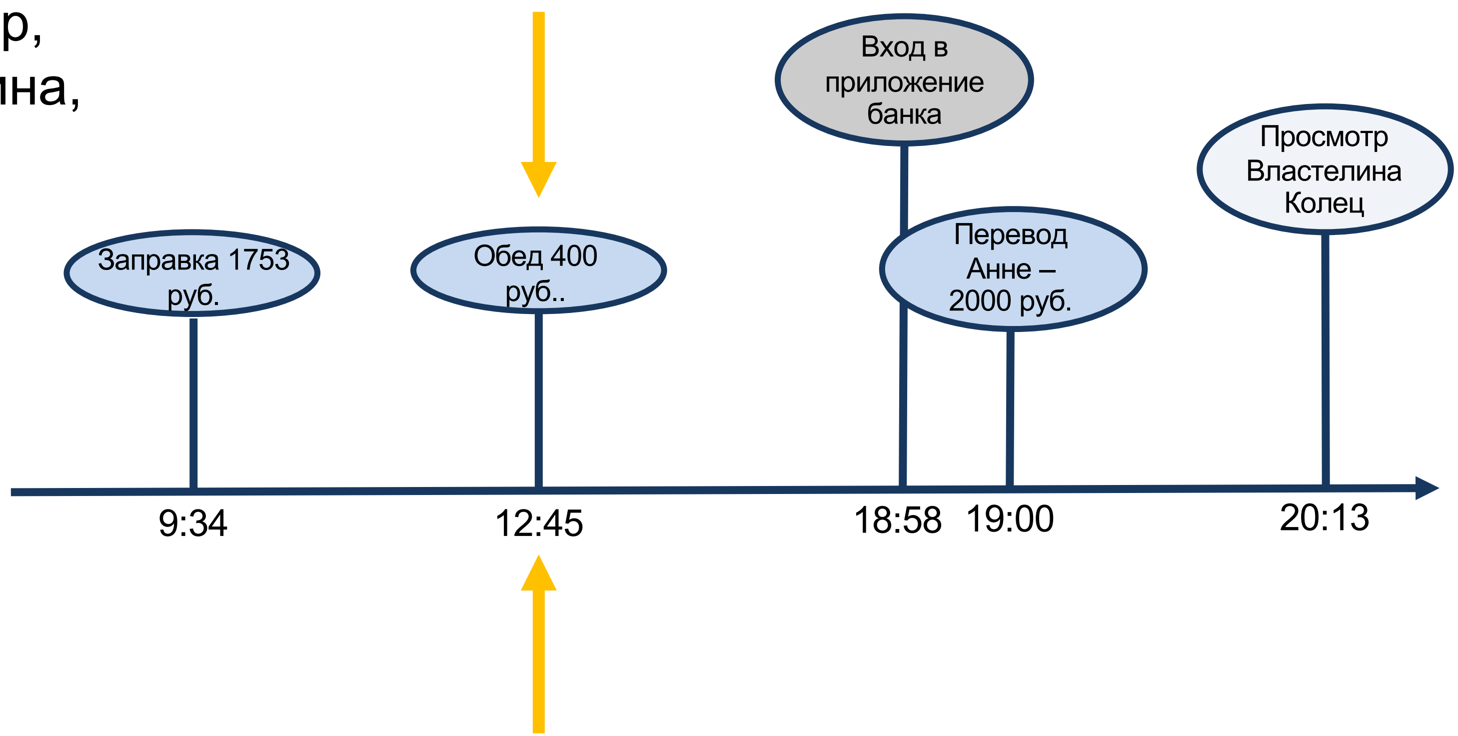
Skoltech

Пример: финансовые транзакции как последовательности событий



Александр,
28, мужчина,
ритейлер

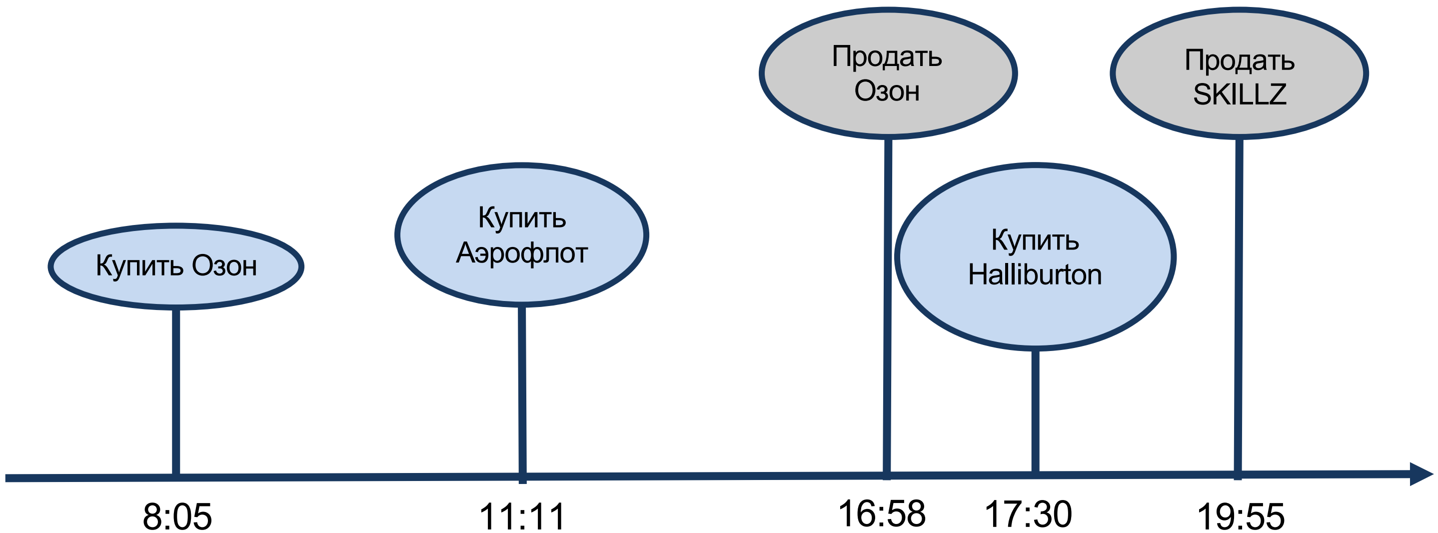
Признаки событий:
сумма, геолокация



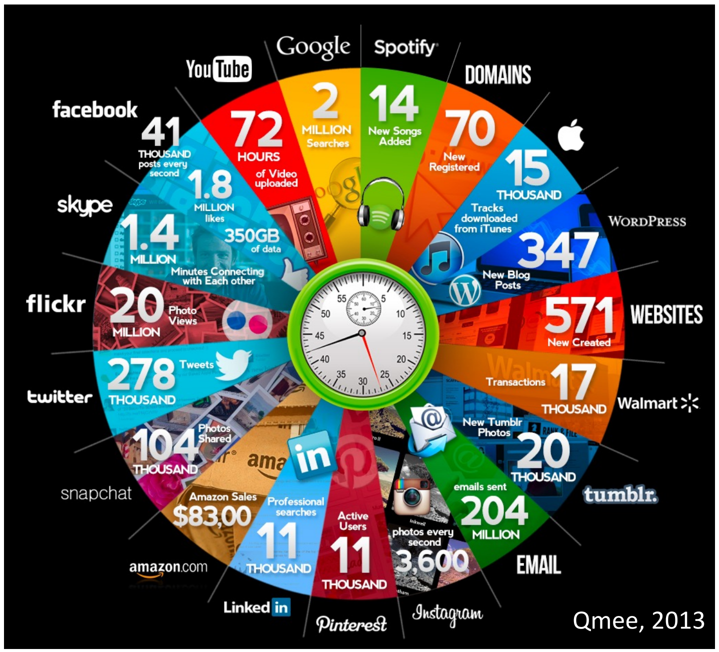
Тип события + момент времени

Пример: операции на бирже как последовательность событий

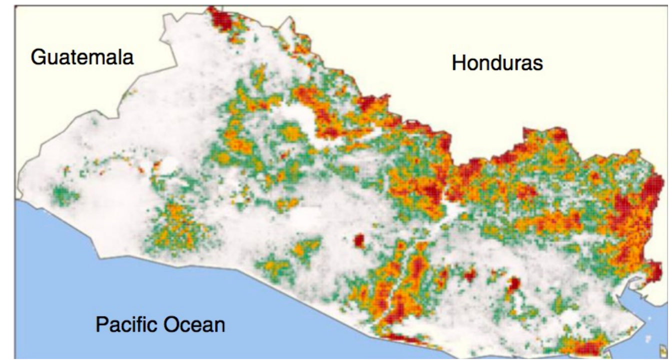
Инвестор



Множество дискретных событий в непрерывном времени



Онлайн активности



Динамика болезней

Финансовые торги

Финансовые транзакции

Разнообразие процессов как причина событий

События – это (шумные) наблюдения за разнообразными сложными динамическими процессами…

Торговля акциями

Распространение гриппа

Создание статей на Википедии

Распространение новостей в Твиттере

Отзывы и продажи на Амазоне

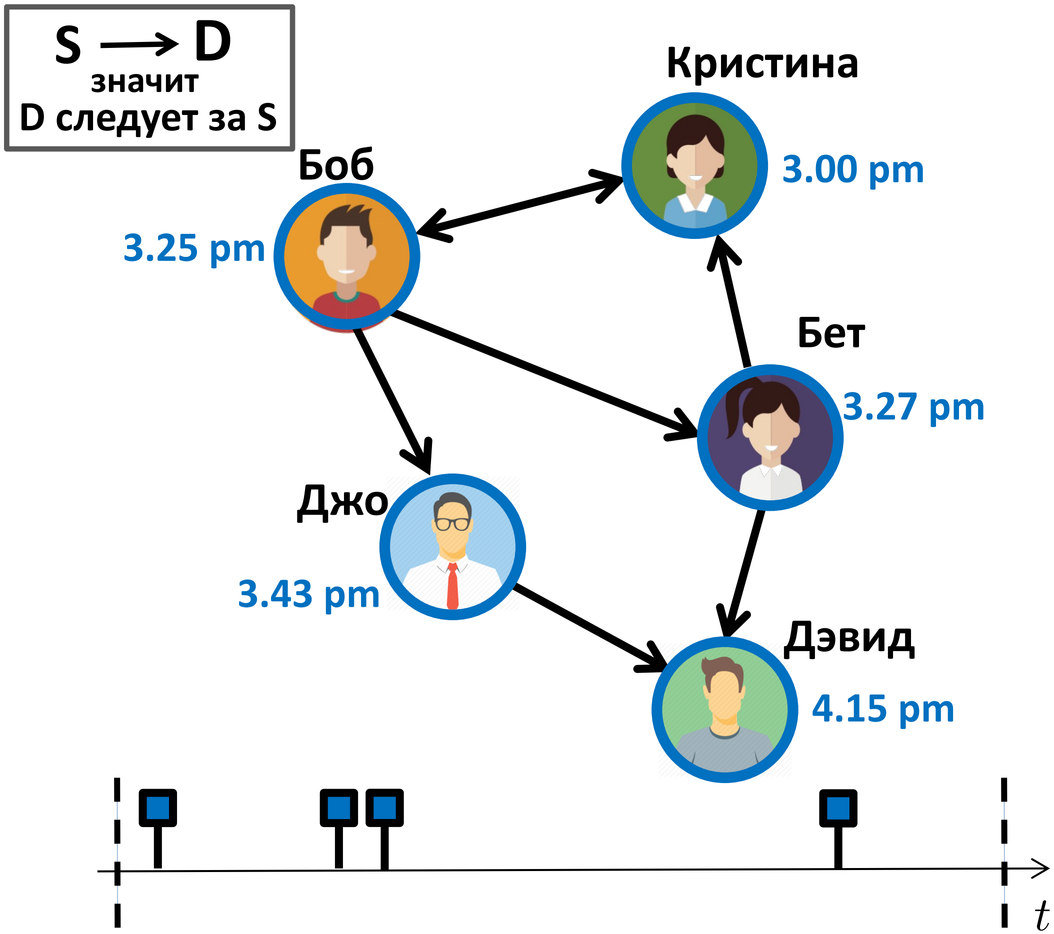
Заказы такси

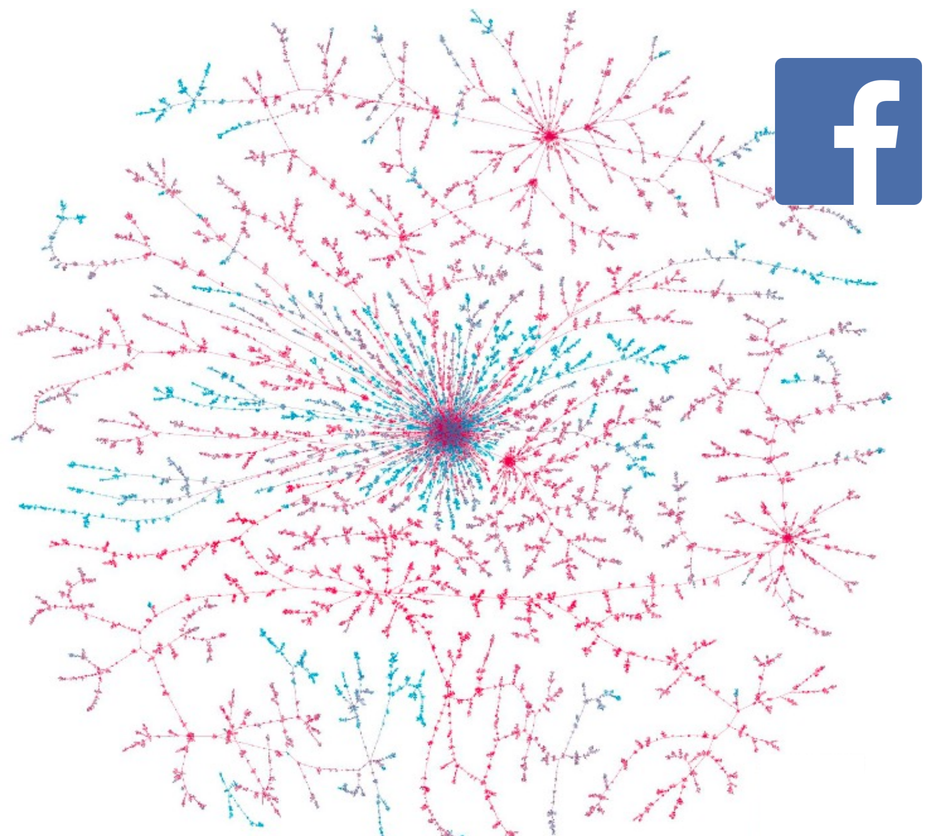
Репутация пользователя в Quora

FAST — SLOW

…в различных временных масштабах.

Пример I: Распространение информации





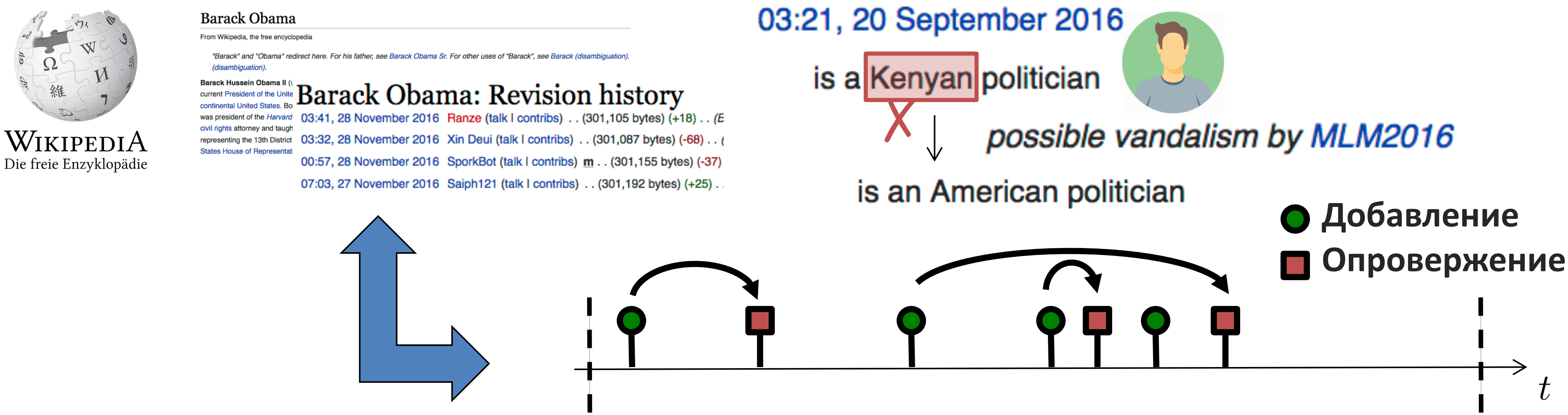
Friggeri et al., 2014

Они могут влиять на события за пределами сети

theguardian

Click and elect: how fake news helped Donald Trump win a real election

Пример 2: история откликов

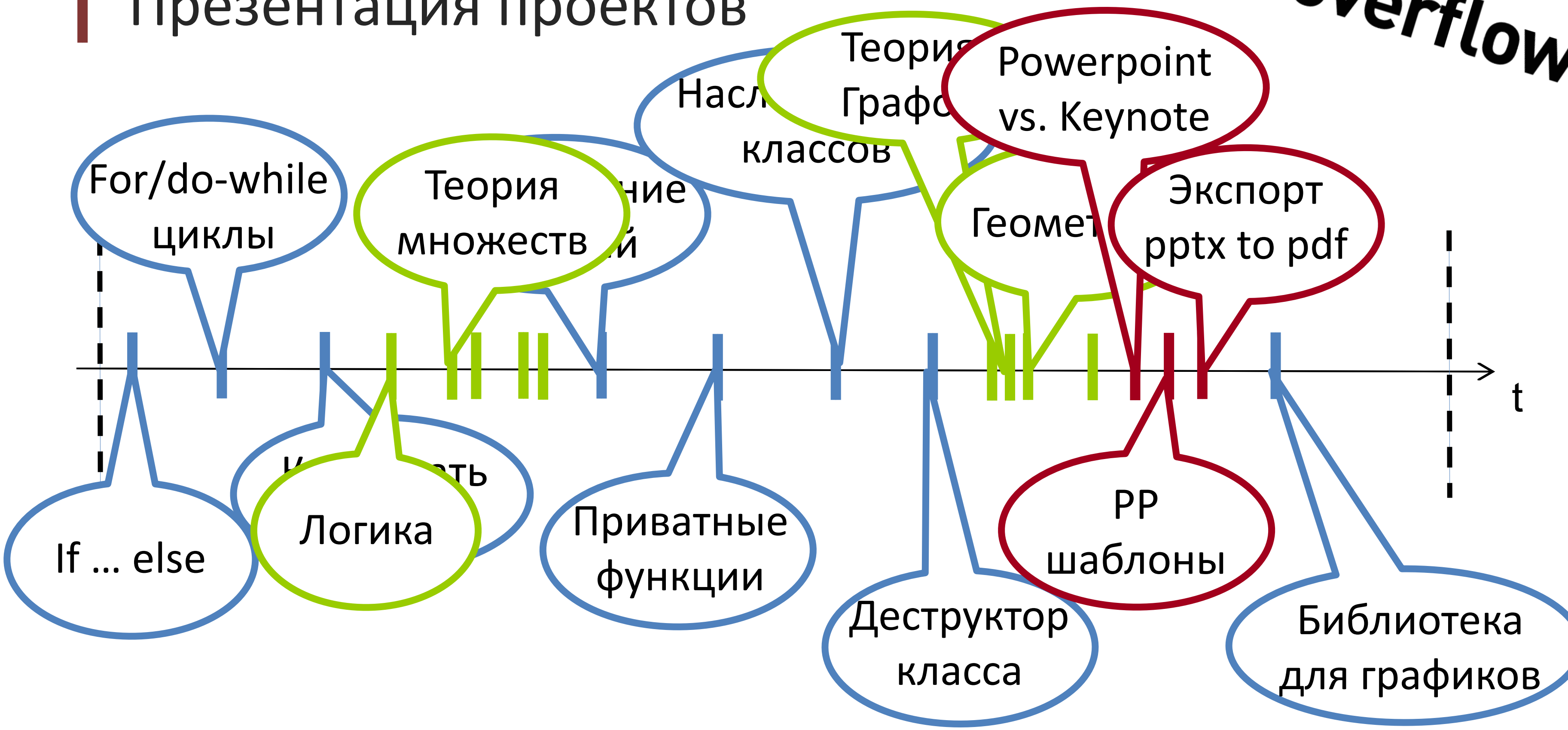


Skoltech

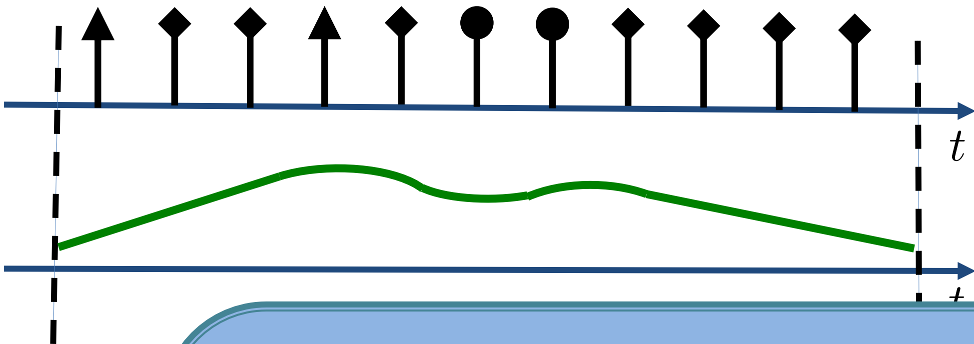
Пример 3: разработка

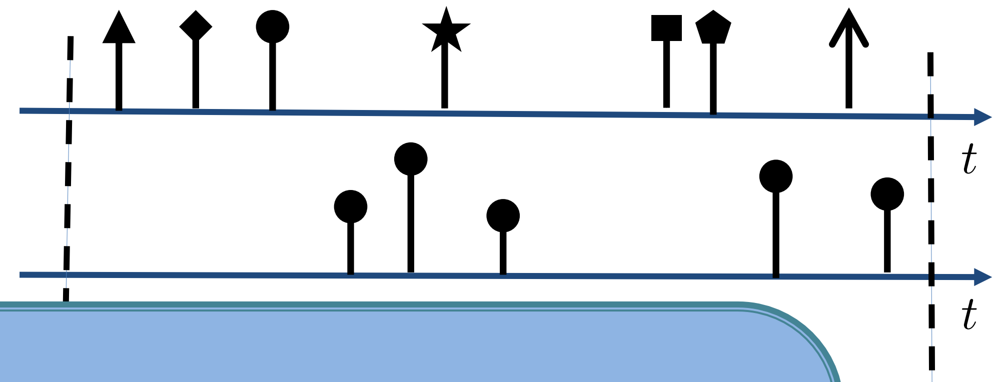
1 год студента по компьютерным наукам

- Введение в программирование
- Дискретная математика
- Презентация проектов



Разве это не просто временные ряды?





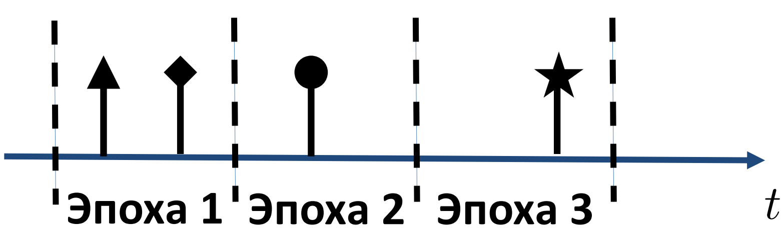
Диск … в … ни

Фреймворк для

временных точечных процессов

дает ***естественное представление***

Чт … с



Что насчет событий зависящих от времени?

Skoltech

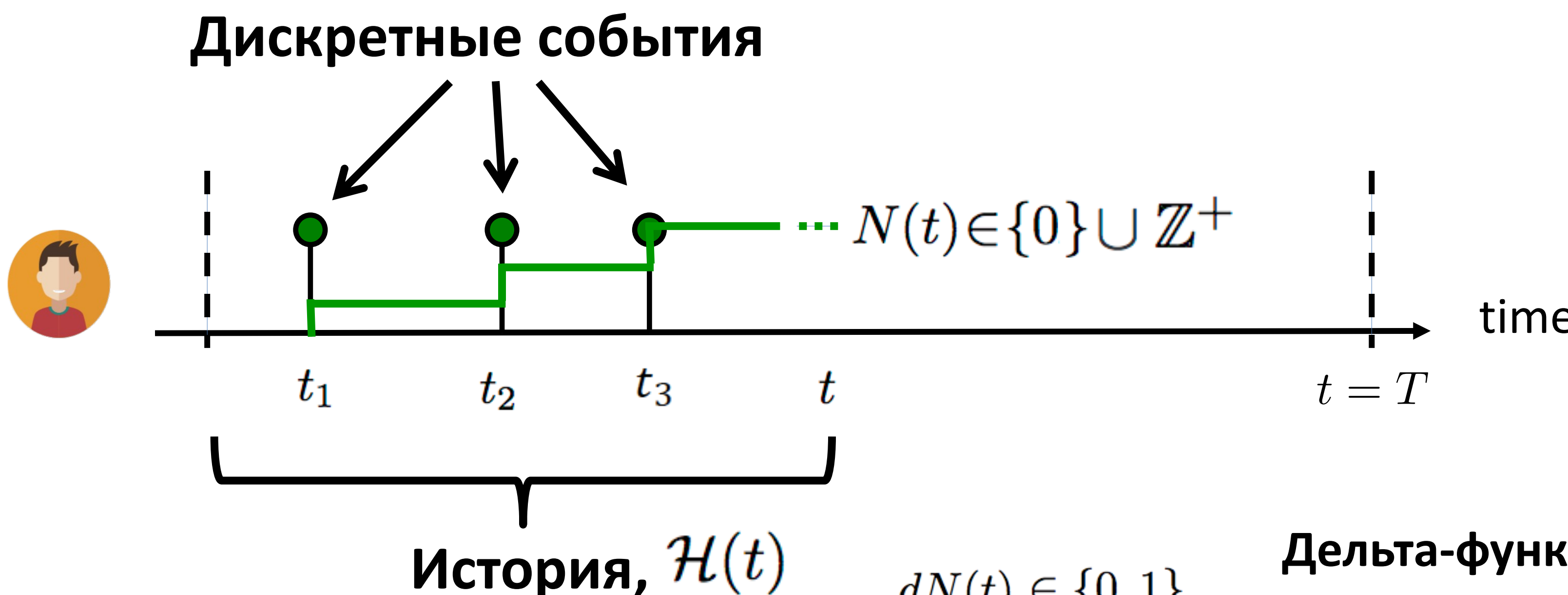
Временные точечные процессы(TPPs): Введение

1. Функция интенсивности
2. Основной строительный блок
3. Суперпозиция

Временные точечные процессы

Временной точечный процесс:

Случайный процесс реализация которого состоит из **дискретных событий локализованных во времени** $\mathcal{H} = \{t_i\}$

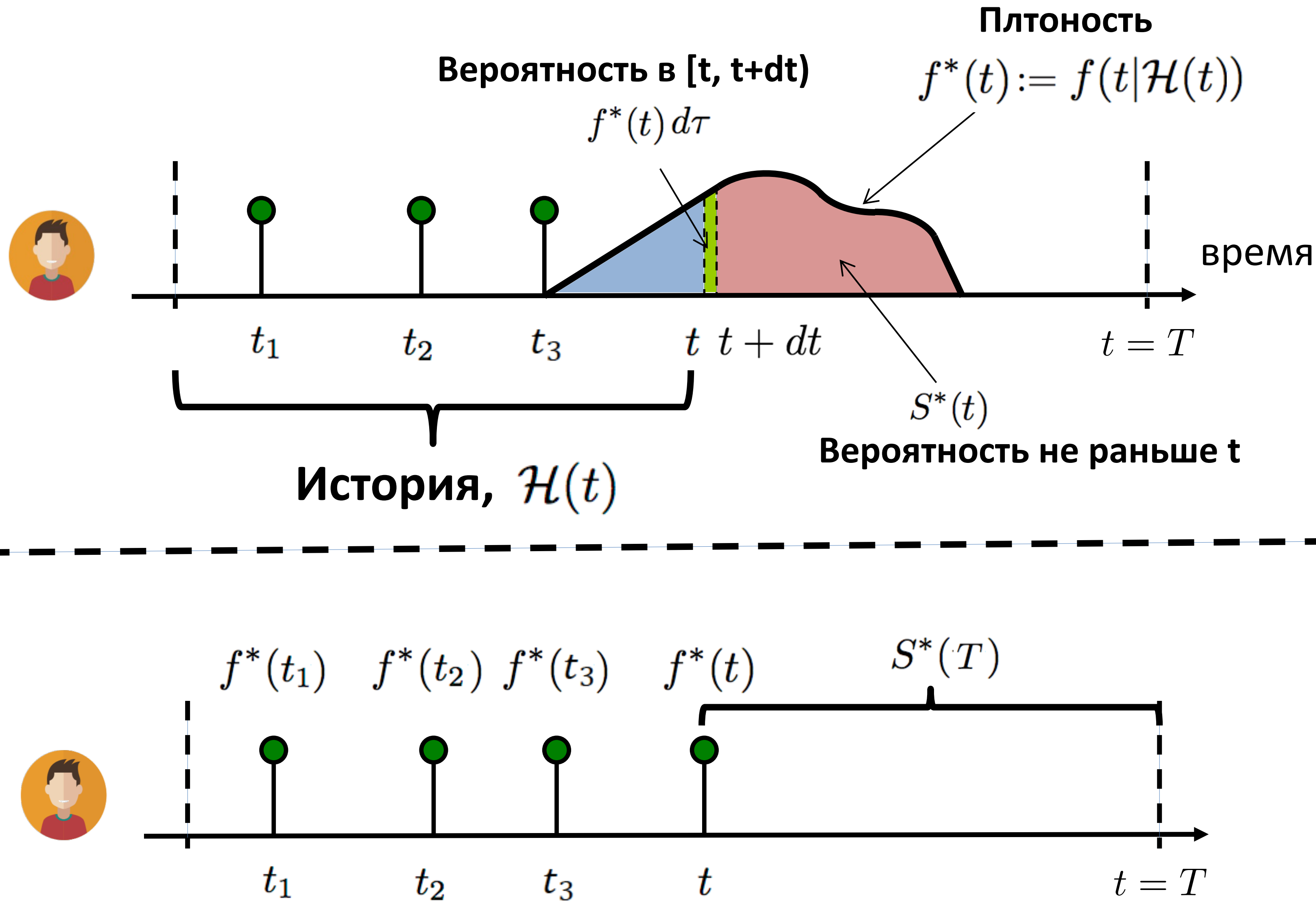


$dN(t) \in \{0, 1\}$

Дельта-функция Дирака

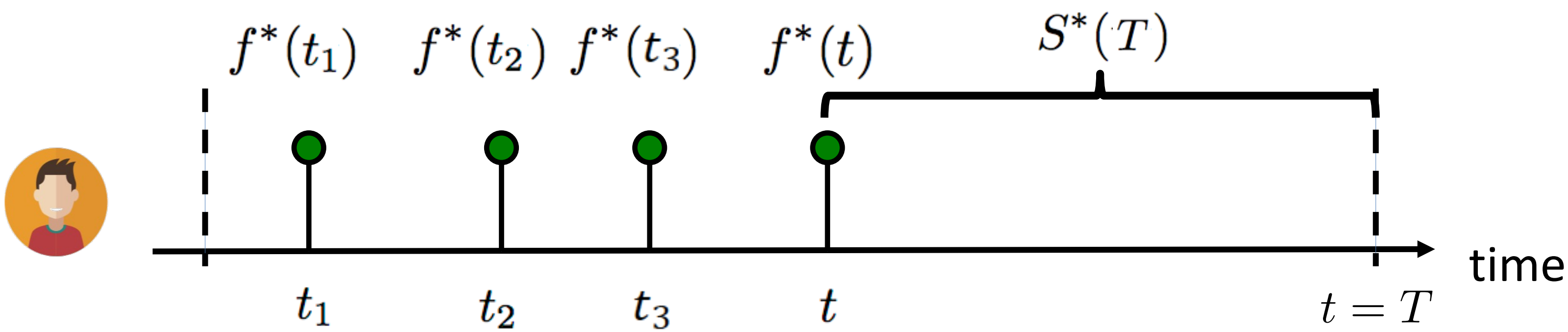
Формально: $N(t) = \int_0^t dN(s)$ ➡ $dN(t) = \sum_{t_i \in \mathcal{H}} \delta(t - t_i)\, dt$

Время как случайная величина



Правдоподобие послед.: $f^*(t_1)\; f^*(t_2)\; f^*(t_3)\; f^*(t)\; S^*(T)$

Проблема параметризации плотности(I)



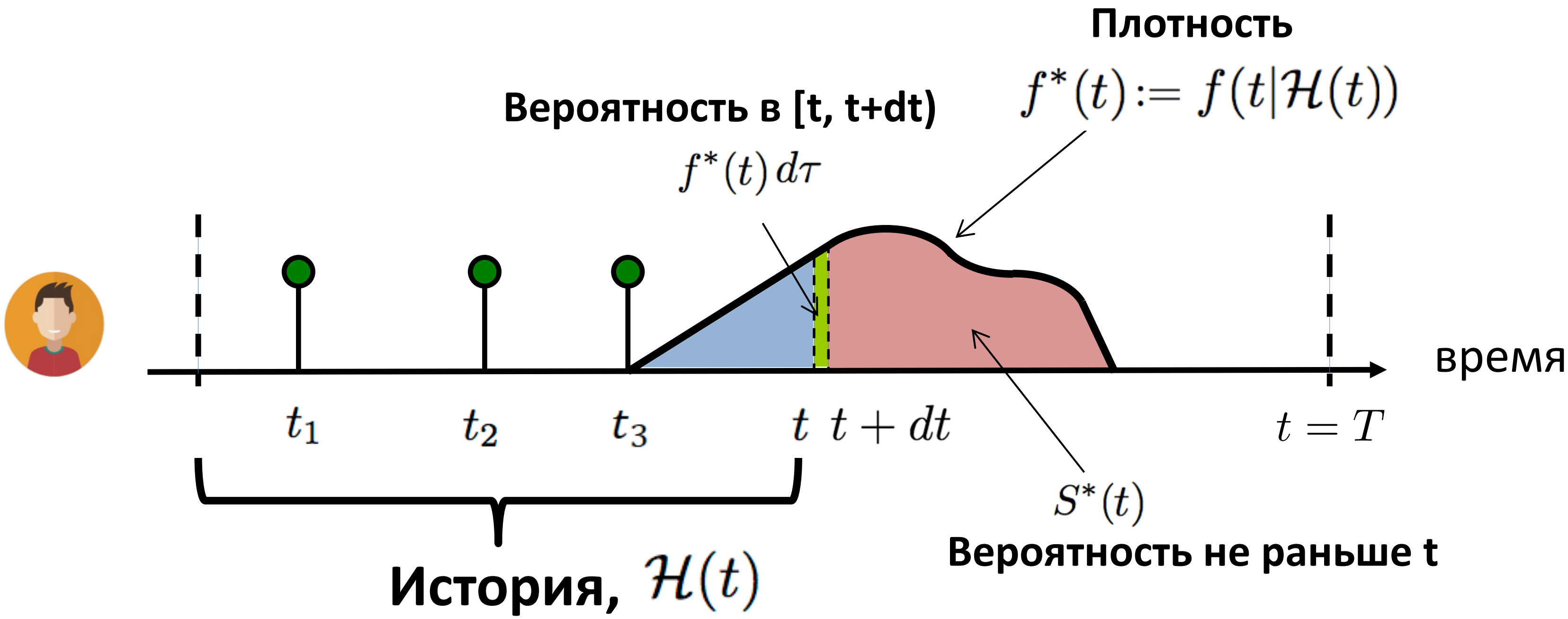
$$f^*(t_1) \quad f^*(t_2) \quad f^*(t_3) \quad f^*(t) \quad S^*(T)$$

$$\frac{\exp\langle w, \psi^*(t_1)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_2)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t_3)\rangle}{Z} \quad \frac{\exp\langle w, \psi^*(t)\rangle}{Z} \quad 1 - \int_t^T \frac{\exp\langle w, \psi^*(\tau)\rangle}{Z} d\tau$$

Сложно для построения модели и интерпретируемости:

1. **Интеграл плотности должен быть равен 1**
2. **Сложно комбинировать последовательности**

Функция интенсивности



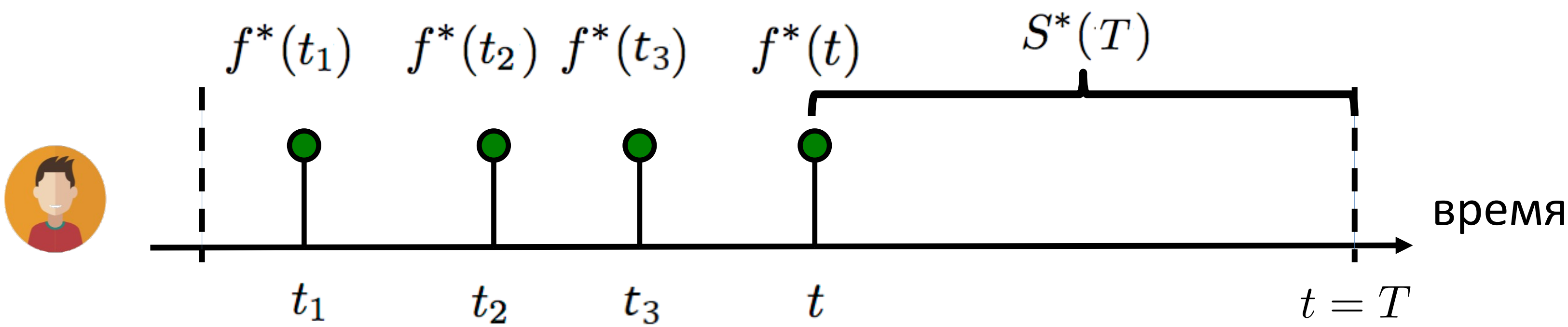
Интенсивность:

Вероятность в [t, t+dt) но не раньше t

$$\lambda^*(t)dt = \frac{f^*(t)dt}{S^*(t)} \geq 0 \quad \Rightarrow \quad \lambda^*(t)dt = \mathbb{E}[dN(t)|\mathcal{H}(t)]$$

Наблюдение: $\lambda^*(t)$ **Частота = # событий / единица времени**

Преимущества параметризации интенсивности (I)



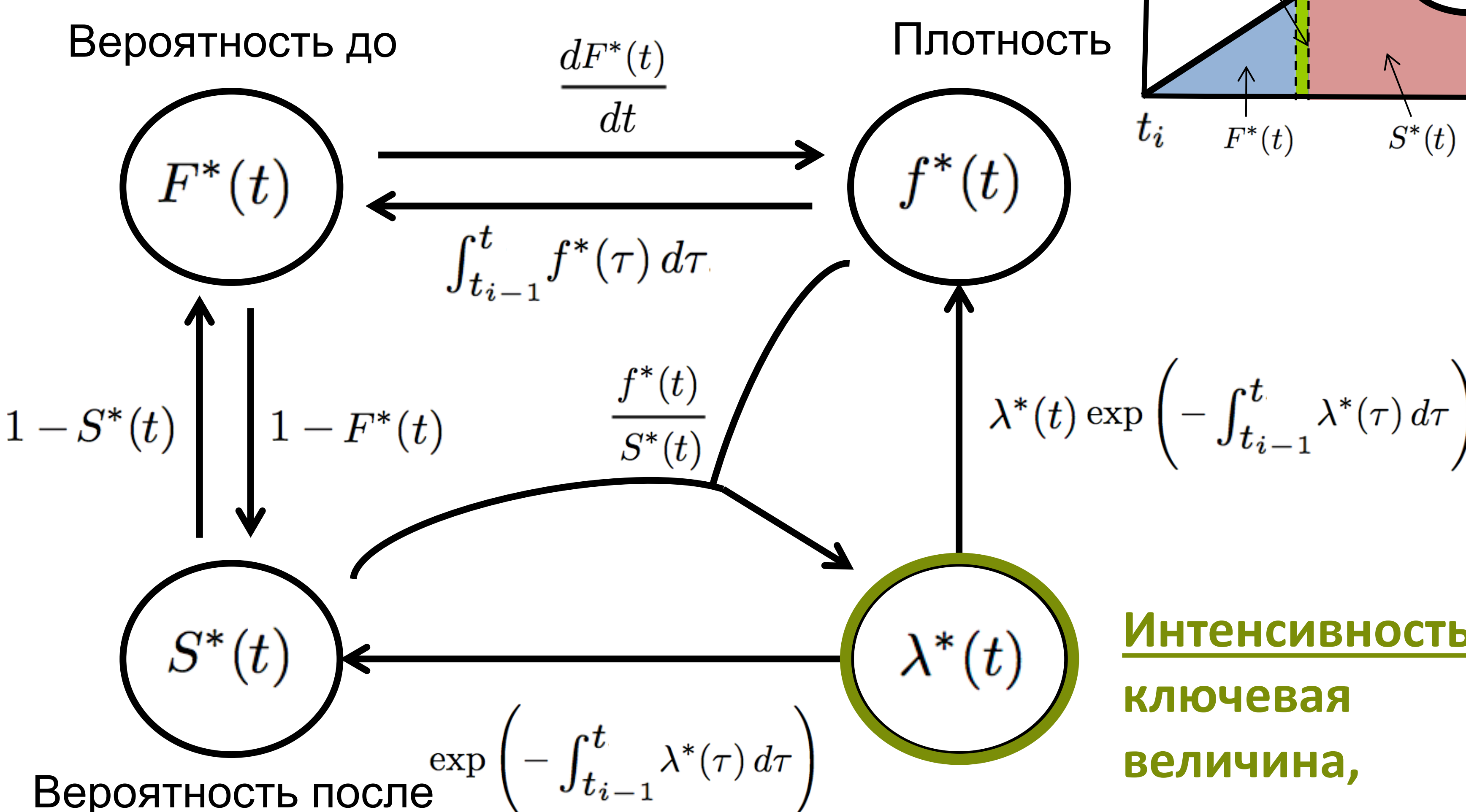
$$\lambda^*(t_1)\ \lambda^*(t_2)\ \lambda^*(t_3)\ \lambda^*(t)\ \exp\left(-\int_0^T \lambda^*(\tau)\, d\tau\right)$$

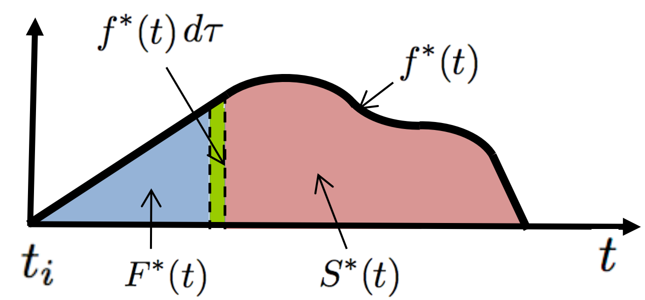
$$\langle w, \phi^*(t_1)\rangle \quad \langle w, \phi^*(t_2)\rangle \quad \langle w, \phi^*(t_3)\rangle \quad \langle w, \phi^*(t)\rangle \quad \exp\left(-\int_0^T \langle w, \phi^*(\tau)\rangle d\tau\right)$$

Подходид для построения модели **и** **интерпретируемости****:**

1. **Интенсивности только неотрицательны**
2. **Легко комбинировать последовательности**

Соотношения между f*, F*, S*, λ*





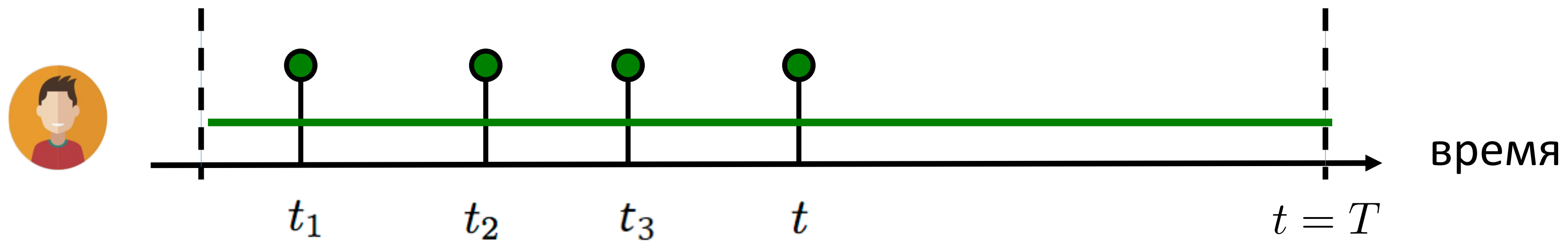
<u>Интенсивность</u> – ключевая величина, которую используют!

Представления:
Временные Точечные Процессы

1. Функция интенсивности
2. Основной строительный блок
3. Суперпозиция

Skoltech

Пуассоновский процесс



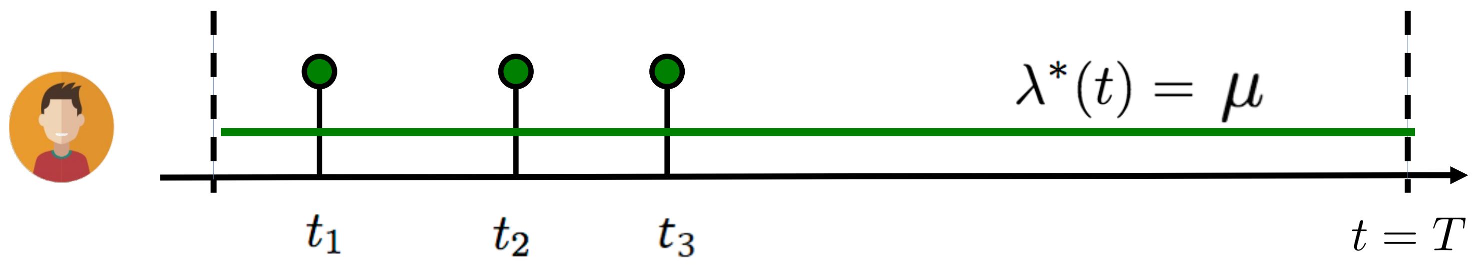
Интенсивность Пуассоновского процесса

$$\lambda^*(t) = \mu$$

Наблюдения:

1. **Интенсивность не зависит от истории**
2. **Однородные случайные появления**
3. **Временные интервалы в соответсвии с экспоненциальным распределением**

Обучение и генерация для Пуассона



Обучение максимизацией правдоподобия:

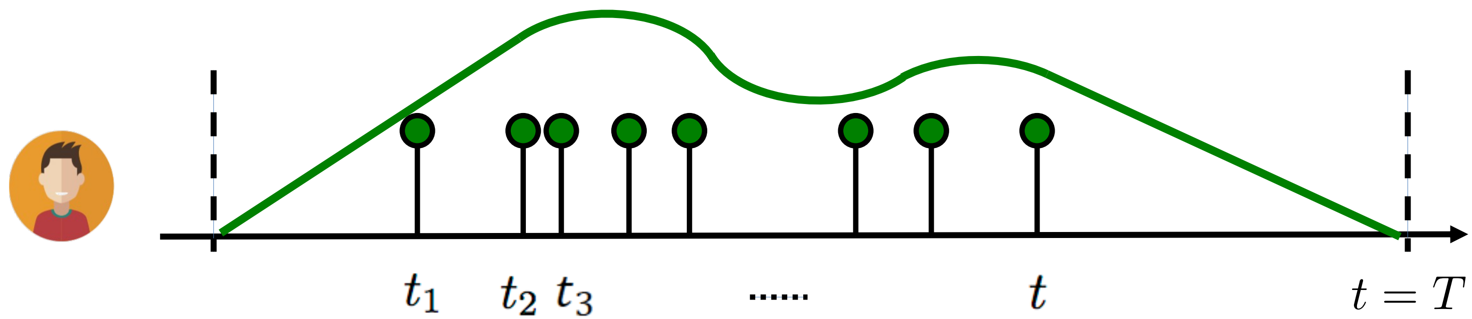
$$\mu^* = \underset{\mu}{\operatorname{argmax}} \; 3 \log \mu - \mu T = \frac{3}{T}$$

Генерация методом обратного преобразования:

$$t \sim \underbrace{\mu \exp(-\mu(t - t_3))}_{f_t^*(t)} \Rightarrow t = \underbrace{-\frac{1}{\mu} \log(1 - \underset{\downarrow}{\overset{Uniform(0,1)}{u}}) + t_3}_{F_t^{-1}(u)}$$

Skoltech

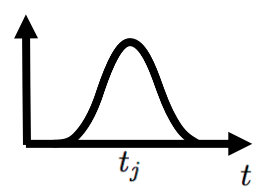
Неоднородный Пуассоновский процесс



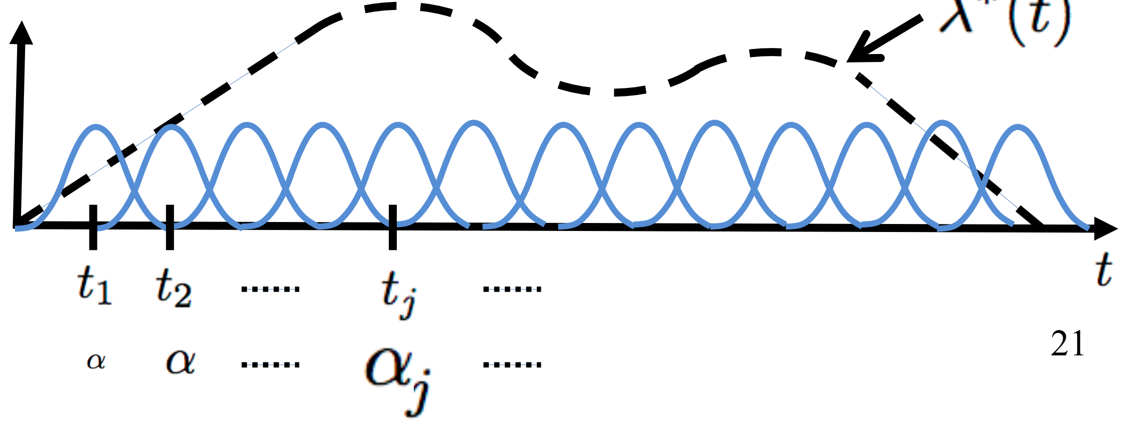
Интенсивность неоднородного Пуассоновского процесса

$$\lambda^*(t) = g(t) \geqslant 0$$ **– Не зависит от истории**

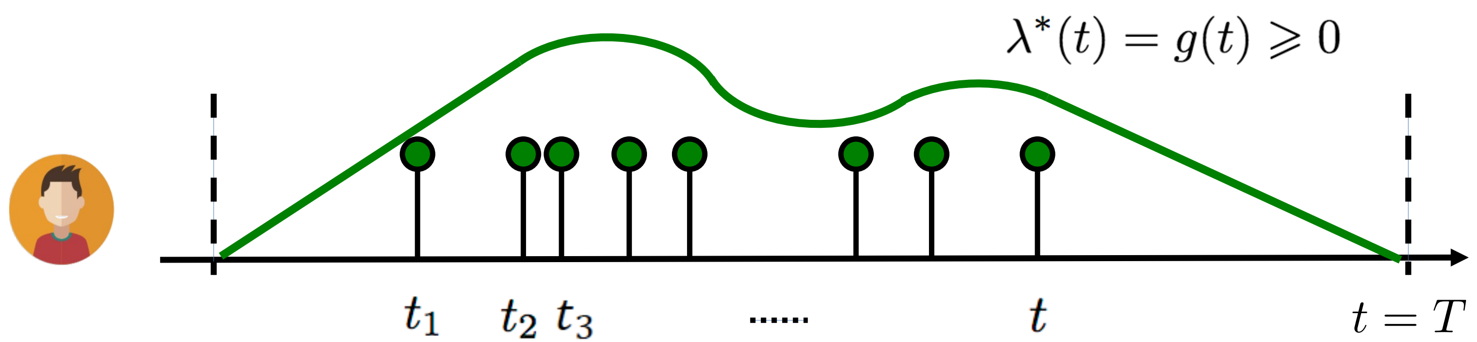
Пример:



$$\lambda^*(t) = \sum_j \alpha_j \overbrace{k(t - t_j)}$$



Обучение и генерация для неоднородного Пуассона



Обучение максимизацией правдоподобия:

$$\underset{g(t)}{\text{maximize}} \sum_{i=1}^{n} \log g(t_i) - \int_0^T g(\tau)\, d\tau$$

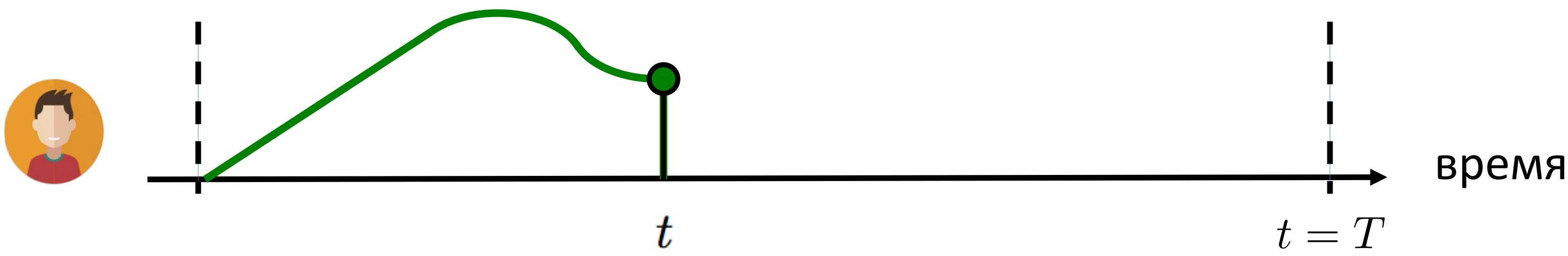
Выборка с отклонением + метод обратного преобразования*:

1. Выбираем t из пуассоновского процесса с интенсивностю μ с помощью метода обратного преобразования
2. Генерируем $u_2 \sim Uniform(0,1)$
3. Сохраняем если $u_2 \leq g(t) / \mu$

$\}$ Сохраняем с вероятн. $g(t)/\mu$

Skoltech

Терминационный (или выживающий) процесс



Интенсивность терминационного (или выживающего) процесса

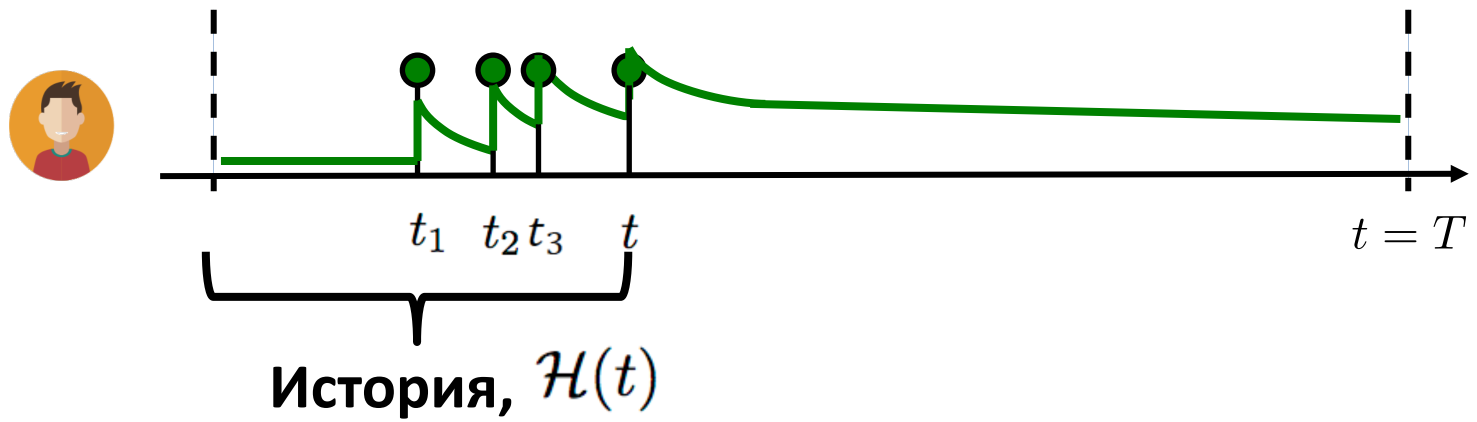
$$\lambda^*(t) = g^*(t)(1 - N(t)) \geqslant 0$$

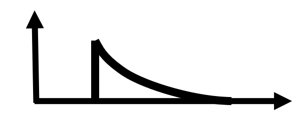
Наблюдение:

1. **Ограниченное количество появлений**

Попробуйте сгенерировать и обучить…

Самовозбуждающийся процесс Хокса





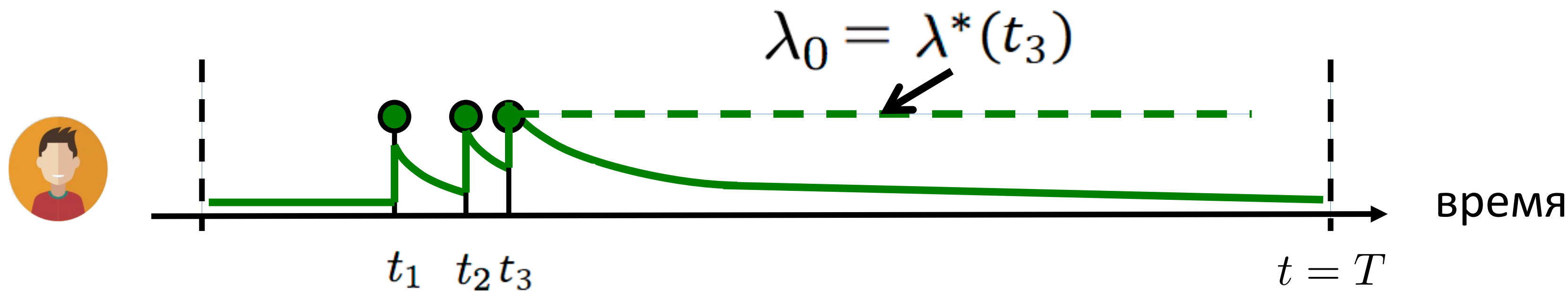
Интенсивность процесса Хокса:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \overbrace{\kappa_\omega(t - t_i)}^{\text{Срабатывающее ядро}}$$
$$= \mu + \alpha \kappa_\omega(t) \star dN(t)$$

Наблюдение:

1. **Кластерное появление событий**
2. **Интенсивность стохастическая и зависит от истории**

Обучение Хоксовской модели



Обучение максимизацией правдоподобия:

$$\underset{\mu,\alpha}{\text{maximize}} \; \sum_{i=1}^{n} \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau)\, d\tau$$

Макс. правдоподбия
совместно выпукло
по μ и α

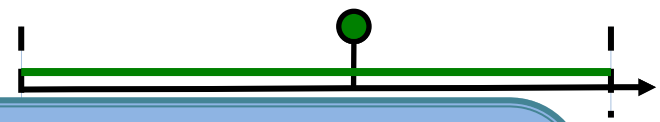
Выборка с отклонением + метод обратного преобразования *:

Ключевая идея: максимум интенсивности λ_0 меняется во времени

Вывод

Строительные блоки представляют **различные динамические процессы:**

Пуассоновский процесс:

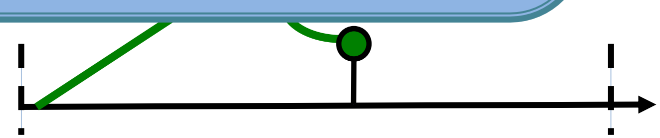


Неод

Мы знаем как обучать их и как генерировать

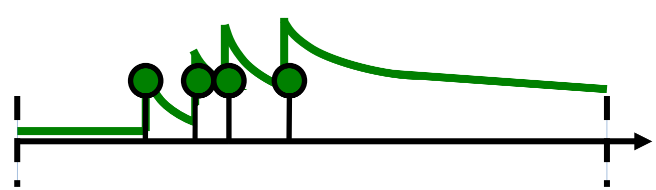
Терм

$$\lambda^*(t) = g^*(t)(1 - N(t))$$



Самовозбуждающийся процесс:

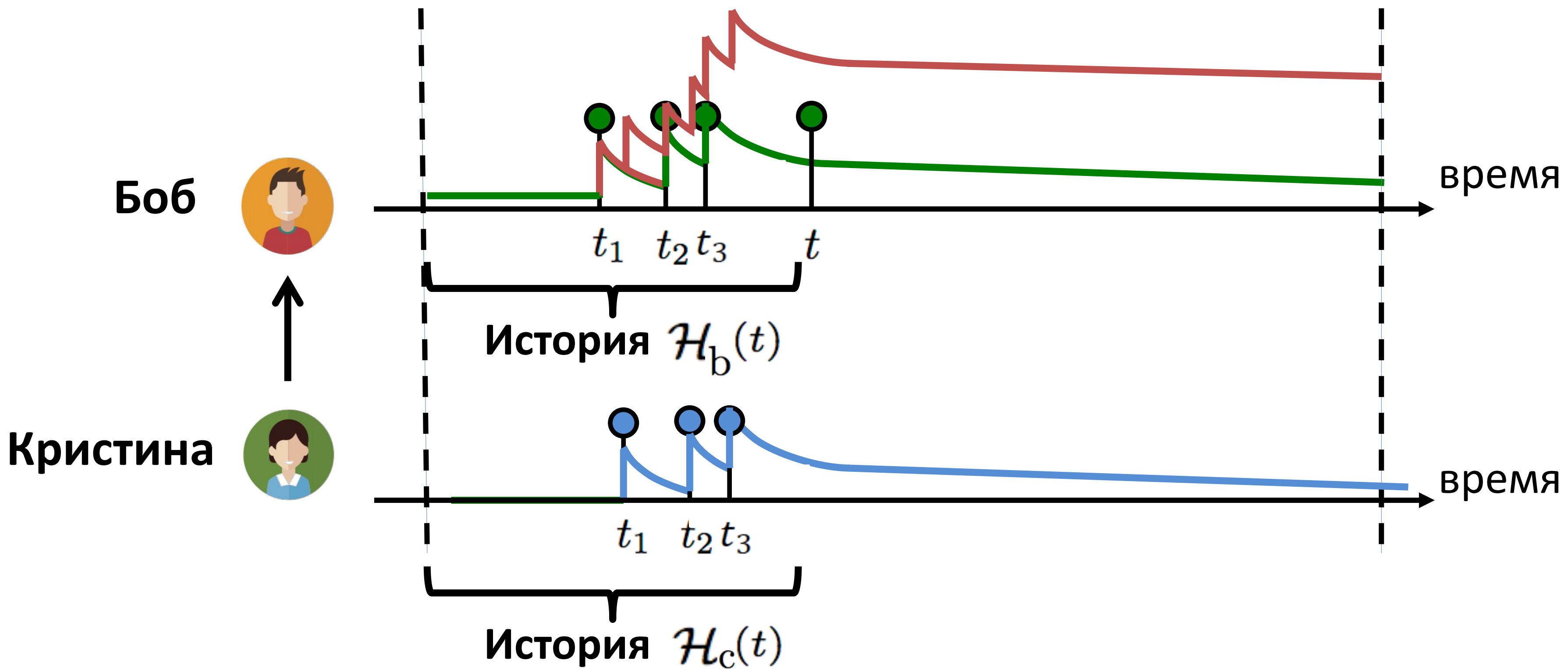
$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$



Представления:
Временные Точечные Процессы

1. Функция интенсивности
2. Основной строительный блок
3. Суперпозиция

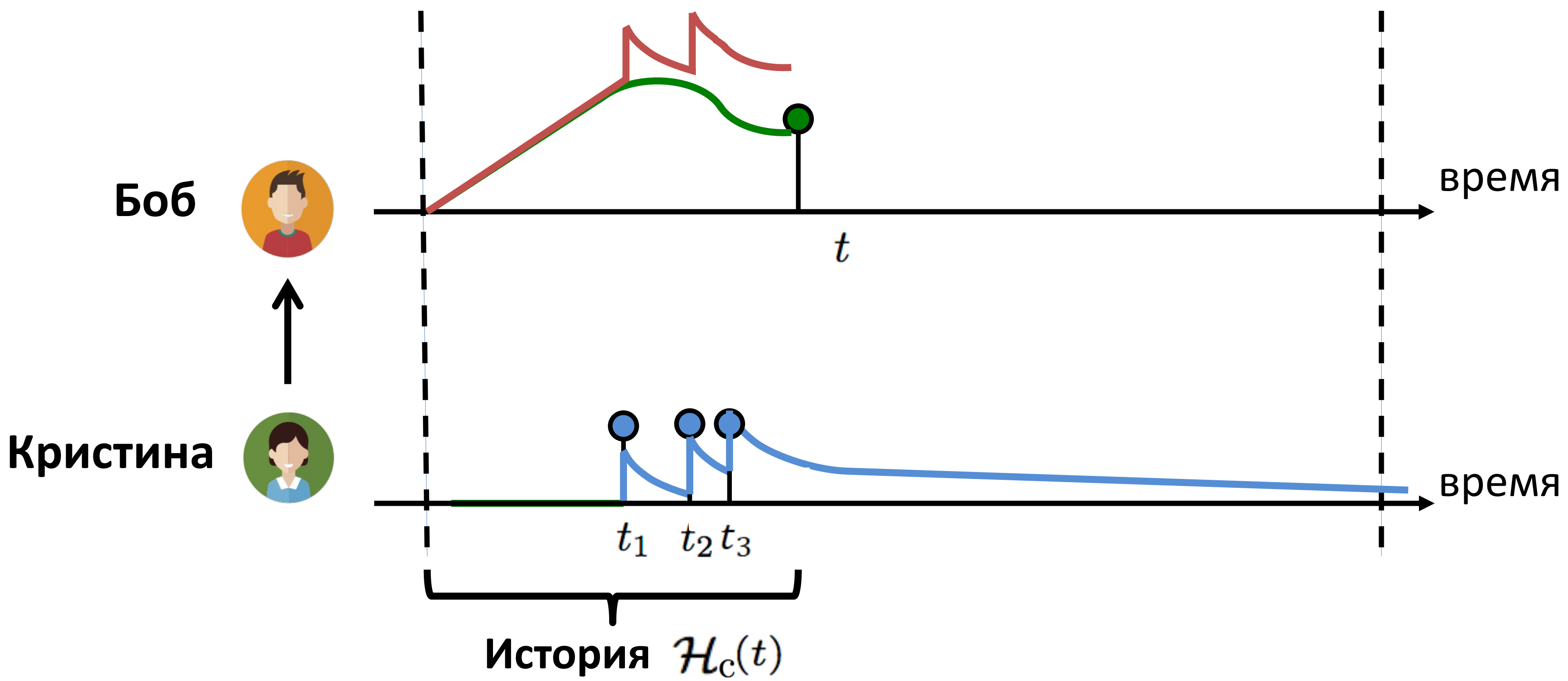
Взаимно возбуждающие процессы



Кластеризованные появления под воздействием соседей

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}_{\mathrm{b}}(t)} \kappa_\omega(t - t_i) + \beta \sum_{t_i \in \mathcal{H}_{\mathrm{c}}(t)} \kappa_\omega(t - t_i)$$

Взаимно возбуждающие терминационные процессы



Кластеризованные появления под воздействием соседей

$$\lambda^*(t) = (1 - N(t))\Big(g(t) + \beta \sum_{t_i \in \mathcal{H}_c(t)} \kappa_\omega(t - t_i)\Big)$$

Skoltech